Statistics Preliminary Examination: May 2011

Instructions:

- Work all 6 problems. Begin each problem on a new sheet of paper, and do not use both sides of the paper (continue on a new sheet of paper if the solution will not fit on a single sheet). Neither calculators nor electronic devices of any kind are allowed. Clearly state any theorem or fact that you use. The 21 parts are approximately equally weighted.
- Abbreviations/Acronyms.
  - pmf (probability mass function); pdf (probability density function); cdf (cumulative distribution function); mgf (moment generating function);
  - MOME (method of moments estimator); MLE (maximum likelihood estimator); UMVUE (uniform minimum variance unbiased estimator); UMP (uniformly most powerful test); LRT (likelihood ratio test); MLR (monotone likelihood ratio).
- Notation.
  - $I(x \in A)$ or $I_A(x)$: indicator function for set A; takes on value 1 if $x \in A$ and 0 otherwise.
  - $\mathbb{E}(X)$: expectation of random variable X.
  - $\mathbb{V}(X)$: variance of random variable X.
  - $X \sim \mathrm{N}(a, b)$: X has a normal distribution with mean a and variance b.
- Common distributions and other results.

  Poisson(λ): $\mathbb{E}(X) = \lambda$, $\mathbb{V}(X) = \lambda$, and pmf

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} I(x \in \{0, 1, \cdots\})$$

  Exponential(λ): $\mathbb{E}(X) = \lambda$, $\mathbb{V}(X) = \lambda^2$, and pdf

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda} I(x > 0)$$

  Beta(α, β): $\mathbb{E}(X) = \alpha/(\alpha+\beta)$, $\mathbb{V}(X) = \alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$, and pdf

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I(0 < x < 1)$$

  Gamma(α, β): $\mathbb{E}(X) = \alpha\beta$, $\mathbb{V}(X) = \alpha\beta^2$, and pdf

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} I(x > 0)$$

  Order Statistics: Let $X_{(1)} \le \cdots \le X_{(n)}$ denote the order statistics from a random sample $X_1, \ldots, X_n$. If X_1 is continuous with pdf $f(x)$ and cdf $F(x)$, the pdf of $X_{(j)}$ and of $X_{(1)} \le \cdots \le X_{(n)}$ is given by:

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!}[F(x)]^{j-1}[1-F(x)]^{n-j} f(x) I(-\infty < x < \infty)$$

$$f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = n! f(x_1) \cdots f(x_n) I(-\infty < x_1 \le \cdots \le x_n < \infty)$$

1. Consider two urns, each containing both white and black balls. Let p_1 and p_2 denote the probabilities of drawing a white ball from the first and second urns, respectively, where $0 < p_1 + p_2 < 2$. Balls are sequentially selected with replacement as follows: with probability α a ball is initially chosen from the first urn, and with probability $1 - \alpha$ it is chosen from the second urn. Subsequent selections are then made according to the rule that whenever a white ball is drawn (and replaced), the next ball is drawn from the same urn; but if a black ball is drawn, the next ball is taken from the other urn. Let U_n denote the event that the nth ball is drawn from the first urn, and define $\alpha_n = P(U_n)$. Similarly, let W_n denote the event that the nth ball drawn is white, and define $\beta_n = P(W_n)$.

   (a) Show that $\alpha_{n+1} = \alpha_n(p_1 + p_2 - 1) + 1 - p_2$, for $n \geq 1$.

   (b) Letting $p = p_1 + p_2$ and $q_2 = 1 - p_2$, use the result of (a) to derive the following expression:

$$\alpha_n = \frac{q_2}{2-p} + \left(\alpha - \frac{q_2}{2-p}\right)(p-1)^{n-1}.$$

   Find $\lim_{n\to\infty} \alpha_n$.

   (c) Derive an expression for β_n, and evaluate $\lim_{n\to\infty} \beta_n$.

2. Let X and Y be continuous random variables, jointly distributed with pdf

$$f_{X,Y}(x, y) = I(0 < x < 1)I(-x < y < x).$$

   (a) Find the marginal pdf $f_X(x)$ of X, and sketch it. Compute $\mathbb{E}(X)$.

   (b) Find the marginal pdf $f_Y(y)$ of Y, and sketch it. Show that $\mathbb{E}(Y) = 0$.

   (c) Find the correlation ρ_{XY} between X and Y.

   (d) Are X and Y independent? Justify your answer.

3. Let the random vector (X, Y, Z) be distributed as,

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \sim \text{trivariate normal} \left(\begin{bmatrix} \mu \\ \mu \\ \mu \end{bmatrix}, \begin{bmatrix} \sigma^2 & \sigma^2/2 & 0 \\ \sigma^2/2 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} \right).$$

   This means in particular that each of X, Y, Z is $\mathrm{N}(\mu, \sigma^2)$, with $\mathrm{Cov}(X, Y) = \sigma^2/2$, and $\mathrm{Cov}(X, Z) = 0 = \mathrm{Cov}(Y, Z)$.

   (a) Show that $X + Y - 2Z$ is independent of $X - Y$, and that $X + Y$ is independent of $X - Y$.

   (b) Find the distribution of:

$$\frac{(X + Y - 2Z)^2}{7\sigma^2} + \frac{(X - Y)^2}{\sigma^2}.$$

   (c) Find the distribution of:

$$\frac{\sqrt{2/3}(X + Y - 2\mu)}{\sqrt{(X - Y)^2 + (Z - \mu)^2}}.$$

4. In the game of *tennis*, two opponents compete to win sets. Suppose player A has probability $0.5 \le p \le 1$ of winning any given set, while player B has probability $q = 1 - p$ of winning any given set. Assume the outcomes of successive sets are independent. In a "best of 3 sets" match, the first player to win 2 sets wins the match. Let the random variable X denote match length, i.e. the total number of sets played in a "best of 3 sets" match. Let $X_1, \ldots, X_n$ be a random sample of n matches drawn from the distribution of X. We are interested in estimating p.

   (a) Find an UMVUE for $\tau(p) = p - p^2$.

   (b) Find the MLE of p.

   (c) Let n_3 be the number of matches in the sample of n that last 3 sets. For $n_3 \le n/2$, what is the relationship between the MLE of $\tau(p) = p - p^2$ and the UMVUE?

5. A random sample $X_1, \ldots, X_n$ is drawn from a Pareto distribution with pdf

$$f(x|\theta) = \frac{\theta \nu^\theta}{x^{\theta+1}} I(\nu \le x < \infty),$$

   where $\nu > 0$ is known and $\theta > 0$ is unknown.

   (a) Find the MLE of θ.

   (b) Show that the LRT of $H_0 : \theta = 1$ vs. $H_1 : \theta \ne 1$ can be based on the statistic

$$T = \sum_{i=1}^{n} \log(X_i/\nu).$$

   (c) Show how to explicitly construct the size α LRT in (b) by using quantiles from a χ^2 distribution.

   (d) Describe an exact (non-asymptotic) procedure to determine a $(1 - \alpha)$ confidence interval for θ.

6. Let X_1, X_2 be a random sample of size 2 from a uniform distribution on the interval $(0, \theta)$. Define $T = X_{(2)}$ and $S = X_1 + X_2$. We seek good tests of $H_0 : \theta \le 1$ vs. $H_1 : \theta > 1$.

   (a) Show that T is a sufficient statistic. Is S also sufficient?

   (b) Consider the test that rejects H_0 if $S \ge 1$. Find the power function, $\beta_1(\theta)$, corresponding to this test, and use it to show that its size is $\alpha = 1/2$. [Hint: consider carefully each of the 3 cases, $0 \le \theta \le 1/2$, $1/2 \le \theta \le 1$, and $\theta \ge 1$.]

   (c) Find the UMP test with rejection region determined by T, whose size is also $\alpha = 1/2$.

   (d) Find $\beta_2(\theta)$, the power function of the UMP test in (c), and show that it coincides with $\beta_1(\theta)$ for $\theta \ge 1$.